SYSTEM AND METHOD FOR INDEXING WEIGHTED-SEQUENCES IN LARGE DATABASES

ABSTRACT OF THE DISCLOSURE

5   The present invention provides an index structure for managing

weighted-sequences in large databases. A weighted-sequence is defined as a

two-dimensional structure in which each element in the sequence is associated with a

weight. A series of network events, for instance, is a weighted-sequence because each

event is associated with a timestamp. Querying a large sequence database by events'

10   occurrence patterns is a first step towards understanding the temporal causal relationships

among the events. The index structure proposed herein enables the efficient retrieval

from the database of all subsequences (contiguous and non-contiguous) that match a

given query sequence both by events and by weights. The index structure also takes into

consideration the nonuniform frequency distribution of events in the sequence data.